DEVELOPING A FRAMEWORK FOR FAKE REVIEW DETECTION USING DEEP LEARNING TECHNIQUES

SHORT SYNOPSIS

submitted in partial fulfillment of the requirement of the degree of

DOCTOR OF PHILOSOPHY

to

Manav Rachna International Institute of Research and Studies
(Deemed to be University)

By
Ms. Richa Gupta
Reg. No.: **21/Ph.D/108**

Under the supervision of

Supervisor

Dr. Indu Kashyap
Professor
Computer Science and Engineering
School of Engineering & Technology
MRIIRS

Joint Supervisor

Dr. Vinita Jindal
Professor
Dept of Computer Science
Keshav Mahavidyalaya
University of Delhi

School of Engineering and Technology
MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH & STUDIES
SECTOR-43, SURAJKUND-DELHI ROAD, FARIDABAD - 121001

SCHOOL OF ENGINEERING AND TECHNOLOGY

Certificate

I hereby declare that the short synopsis "**Developing a framework for fake review detection using deep learning techniques**", being submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy in Computer Science under the school of Engineering and Technology of Manav Rachna International Institute of Research & Studies, Faridabad during the academic year 2022-23, is a bona fide record of original work carried out under the supervision of **Dr. Indu Kashyap**, Professor, FET, MRIIRS and co-supervision of **Dr. Vinita Jindal**, Professor, Department of Computer Science, Keshav Mahavidyalaya, and has not been submitted in full or part to any other university or institute for degree or diploma. Further, it is certified that this short synopsis is plagiarism free, and the proof has been submitted.

23/11/23
(Ms. Richa Gupta)
Reg. No. 21/Ph.D/108

23/11/23
(Dr. Indu Kashyap)
Professor
Faculty of Engineering & Technology
MRIIRS

23/11/23
(Dr. Vinita Jindal)
Professor
Keshav Mahavidyalaya
University of Delhi

23/11/23
(Chairperson, DRC)

23/11/23
(Director, CDP)

ABSTRACT

Reviews on the web, tend to sway the decision of prospective consumers, who are planning to buy a particular product or service. If there are many positive reviews of a product, the customer feels inclined to buy that product even if it is costlier than other products in the same category. However, if many reviews are negative, the customer is wary of buying it. This promoted the fake review creation business. Many companies buy fake reviews in bulk, which can be used to downgrade their competitors or artificially improve their credibility. Fake reviews may be generated by a machine or by people who have been paid for them. Fake reviews are near to impossible to be detected by the human eye. Hence, the need for fake review detection is gaining popularity.

Most of the researchers have worked on machine learning and deep learning models. But despite notable progress, several research gaps persist such as the small size of training data, little or no consistency of solution on different datasets, concept drift, the adaptation of fake reviews over a period of time, multilingual, multimodal challenges, and imbalanced datasets.

All these factors served as the basis to pursue research work in the area of fake reviews and aim to look into more robust solutions using various AI models. Various deep learning models including transformers and language generators will be explored to solve the problem of imbalance. Along with this, the problem of insufficient data in unexplored domains will be investigated to find an optimized solution for it. In conclusion, fake review detection is a dynamic and evolving field that requires continuous research efforts.

Keywords: **Deep Learning, Fake Review Detection, Machine Learning, Online Reviews**

Contents

1. INTRODUCTION 1
2. LITERATURE REVIEW 4
3 RESEARCH GAPS 12
2 PROBLEM DEFINITION 14
3 OBJECTIVES OF THIS RESEARCH 14
4 RESEARCH METHODOLOGY 14
5 EXPECTED OUTCOMES OF THIS RESEARCH 17
References 18
Appendix 23

1. INTRODUCTION

Online reviews are dominating e-commerce websites today. From buying products from Amazon and/or Myntra to hotel reviews on TripAdvisor to restaurant reviews on yelp, to service reviews on Google, everywhere the consumer/user reviews help shape the business/service domain. These reviews, if genuine, help the service provider/seller to improvise their product or service. But, if these are fake, they mislead the service provider as well as the future customers who may/may not buy the product upon reading the reviews. These fake reviews are a form of disinformation and misinformation. Dis-information means giving fraudulent reviews to harm others, whereas misinformation means simply giving false information without any wrong intentions. Either way, these fake reviews are harmful to society on the whole. The internet has already created a space to falsely push narratives and opinions for self-benefits that may harm others [1]. All parties, including consumers, business and service providers, and e-commerce platforms, are at a disadvantage if they can't identify fake reviews. Consumer needs to be protected against these reviews. Businesses and service providers should be defended against artificially created competition and their reputation should be protected, as reviews affect product ranking [2]. The e-commerce and online travel platforms should be able to shoulder the responsibility of the review channels in a fair manner. Giants like Google and Yelp already employ fake review detection algorithms, but these are not available in the public domain. Amazon keeps on deleting fake reviews over time with an average lag of 100 days. But it might happen that a large number of consumers have already seen the reviews and the damage has been done before the reviews are deleted.

Fake reviews may be positive or negative. The motive for giving fake positive reviews might be monetary rewards or incentives in the form of discounts. They may be for helping a friend at the start of a business or just for enhancing self-esteem [3]. The negative fake review may be motivated by monetary compensation or as a weapon for revenge/betrayal. The flood of negative reviews against a brand can lead to unfair competition as well as lowering its rank on online platforms such as Google, Amazon, TripAdvisor, etc. [4]. The

authors Zaman et al. [3] interviewed 20 South Korean and 31 French customers of a cosmetic brand, who have given fake reviews to compare them cross-culturally. However, that research suffered from the lack of comparison between sectors. The motives mentioned were particularly applicable to buying products and not to the tourism or hospitality industry. Also, only the motives of consumers had been included. The point of view of business owners or brand managers can also be an interesting problem. Abdulqader et al. [5] feel that deception theories in psychology may be the key to opening new areas for fake review detection. These theories study human behavior as faking reviews is deception only. Ten theories have been used, like the Leakage theory, four-factor theory etc. to identify the fraud reviews. The important deceptive elements were gathered from these most popular deception theories and a theoretical model was formulated. This model was evaluated using classification methods (Logistic Regression, Naïve Bayes, Decision Tree, and Random Forest) on preprocessed datasets.

Many brand marketers and service providers engage in unscrupulous behavior and buy fake positive reviews from known or unknown reviewers. This is done to give a 'cold start' to new products or to increase the popularity of brand by artificial inflation of favoring reviews. Taking the help of fake negative reviews to artificially harm the reputation of competitors is a common problem. He, Hollenbeck and Prosperio [2] tracked the relationship between product reviews, average ratings, prices, and sales rank on Amazon.com and established direct evidence of detailed firm outcomes from buying fake reviews.

False information, be it in the form of news or reviews or just tweets on social media, disrupt society and cause social troubles. For example, false COVID--19-related news on vaccinations and lockdowns on social media triggered bad behavior at large [6].

Humans rely heavily on online reviews. A survey from Brightlocal 2019, shows that 76% of consumers trusted these reviews in the way they trust recommendations from friends [7]. Despite the advisories issued by the government and websites time-to-time, it is very difficult for an ordinary consumer to detect fake reviews accurately. However, some features can be made available to them, so that they can make better decisions regarding this [8]. The work done by Kostromitina et al. [9] analyzed Yelp reviews and found why

restaurants are given extreme star ratings and what are the characteristics customers look for. But the number of fake reviews that may be there, was not considered.

Fake review detection can be treated as a Natural Language Processing (NLP) problem where text can undergo machine learning models or long-short memory networks to find out whether the review written is authentic or false. Many supervised, semi-supervised, and unsupervised models have been used to find the efficiency of various models based on datasets from Yelp, TripAdvisor, Amazon, YouTube, etc. But the fraud reviewers keep on adding new features to their reviews to avoid being caught by the current approaches [10] [11] [12].

Companies like Amazon suddenly seeing a spike in no. of unverified reviewers with 5-star ratings, is a sign of fake reviews according to Abdulaqader et.al. [5] . Hence, online platforms like Amazon, Yelp, and Google should keep updating their fake review detection strategies [4]. These large number of fake reviews affect the credibility of the brand or product in consumers' minds [13].

According to Salminen et al. [4], there are two types of fake reviews. One that is human-generated, which has been the focus of discussion till now. Second is the computer-generated review where the machine automatically creates a fake review. The latter are cheaper in terms of money and time, for businesses. Technology is reaching such heights that humans can't make out the author even if it is authored by an actual individual [14].

The authors Salminen et al. [4] generate fake reviews using the Amazon dataset (2018) and create two language models ULMFiT and GPT-2. This dataset comprises 40,000 samples for training purposes and an additional 10,000 sample for testing, spanning across diverse categories. The advantage of creating this is that the dataset becomes equally balanced. The authors evaluate these utilizing both quantitative and qualitative assessment to select the superior language model, which in this case is GPT-2. Then a classifier is applied to detect these synthetic fake reviews from authentic ones and the same exercise is given to recruited human workers. The results are compared, and findings indicate that machines give more accurate results than humans.

2. LITERATURE REVIEW

Many researchers have tried to solve this problem of identifying fraud reviews using various machine learning and deep learning models. In this section, these studies have been discussed based on the models used and the identifying features. This literature review aims to provide an in-depth exploration in the field of fake review detection, offering insights into the motivations, methodologies and advancements that have shaped this dynamic and evolving domain. This review seeks to highlight the key developments and research gaps that serve as the basis for finding efficient fraud review detection mechanisms.

Various algorithms including rule-based, graph-based, machine learning, and deep learning have been used to classify fake and real reviews. The most recent ones are mentioned in this section. Supervised, semi-supervised, and unsupervised machine learning models have been frequently employed by researchers in the field of fake review identification.

Support vector machine, K-Nearest Neighbor, and Linear Regression etc. are the most common supervised machine learning models. All these are used by Tufail et al. [15] to detect fake reviews on the Yelp hotel review dataset. The model proved to be a robust one, but it only focused on supervised models. Similarly, The authors Poonguzhali et al. [16] have implemented SVM for fake review detection by creating an online e-commerce interface. Fake reviews are predicted in this system and are not used in the database while recommending a product on this interface. But this method can't upload more than one review by one user. So, the prevention method is not feasible in large databases. Hence, Alsubari et al. [17] have extracted features such as sentiment score, four grams, number of verbs, nouns, and strong positive or negative words to identify fake reviews. Then four different supervised classifiers namely, SVM, Naïve Bayes, Random Forest, and Adaptive Boost are applied, and their accuracy compared for fraudulent reviews detection. The limitation of this study was fewer extracted features and, the dataset was limited to the hotel domain. The work done by Abdulqader et al. [5] observed that non-verbal features carry greater significance than verbal features and the amalgamation of both can enhance

the detection accuracy. But only a pure theory-based model has been given in this work, which may or may not apply to datasets other than the one used here.

The supervised machine models use extensive labeling, which is laborious as well as subjective. Hence, the semi-supervised approach uses a pre-defined set of features to train classifiers. Jing-Yu and Ya-Jun [18] employed a semi-supervised model known as AspamGAN (Generative Adversarial Network) which identifies fake reviews through the incorporation of an attention system within the classifier. Data from TripAdvisor has been used. But, if the data generated by this model is insufficient, it may result in poor accuracy. The earlier version of this method spamGAN given by Stanton and Irissappane [19] was a powerful tool that used a simple classifier. But it had the disadvantage that it may lose important information such as context, the focus of the sentence, etc., and thus won't be able to detect fake information with high probability. AspamGAN has the advantage that it has better performance with limited label data, than SpamGAN.

As mentioned earlier, detecting fake reviews accurately is highly improbable by humans. Hence, the availability of labeled data is less. Unsupervised models have a good scope here. The work done by Mothukuri et al. [20] creates clusters using the extracted features via K-means clustering, GMM Full covariance clustering, and GMM Diagonal covariance unsupervised model to detect fake reviews taken from the café dataset of Yelp. It was found that K-means shows the highest accuracy among the three. Many other unsupervised models could have been explored in the aforementioned work, and in different domains.

There is a major problem of concept drift [21], while using machine learning models. Concept drift refers to the adaptation of fake reviewers over a period of time. They start writing in such a way that their writing skills are similar to real reviews and hence, can't be detected by the detection algorithms in place. For this, Wang et al. [22] have incorporated temporal patterns in reviews. A suspicion degree is added for fake review detection based on 3D time series, the number of reviews, and the information entropy. This seems comprehensive, but only supervised models have been used. One major drawback of machine learning models is that they require a manual process for extracting features and a substantial amount of computational resources [23].

Kumaran et al. [24] have also used naïve bayes, Eristic regression, and SVM for detecting fake reviews in a dictionary based on social media keywords and online reviews. This model uses a very limited feature set of unigrams, bigrams, and length of review. Along similar lines, Hussain et al. [25] have given some important behavioral and linguistic pointers to identify fake reviews. For example, the absence of profile data of the reviewer, posting duplicate reviews, short and often grammatically erroneous reviews, groups of reviews with the same timestamp, and excessive use of positive or negative words. But the behavioral model and linguistic model have been studied separately. The chances of reviews being classified accurately decrease when only one model is used.

Many supervised, semi-supervised and unsupervised models have been used to identify fake reviews, to date. These traditional machine learning models perform adequately on small datasets and are highly valued by researchers. In addition to this, they are simpler to implement and computationally cheap but, their performance deteriorates on larger datasets as compared to deep learning models. Also, their solution doesn't sustain over time and the problem of concept drift arises. In addition to the above shortcomings, they also require manual feature extraction and huge computational resources. Hence, deep learning models were explored by researchers.

Deep learning models such as Convolutional Neural Networks (CNN), Recurrent CNN, Long-Short term memory (LSTM), Bidirectional Gated Recurrent Unit (Bi-GRU) with attention, Generative Adversarial Networks (GANs), and Bidirectional Encoder Representations from Transformers (BERT), have all been used to detect fake reviews. These models use neural network architecture with multiple layers, similar to the human brain, and thus can accomplish better solutions to big data problems. Hence, they are getting popular in solving fake review detection problems.

The authors Ren et al. [26] use graph neural networks to find associations between reviews, users, and products. This is based on the premise that there is a dependency between the user and product, the reviewer and the time, ratings, etc. Then introduce the idea of suspicious values based on the TrustRank [27] method is introduced. This model finds more fake reviews, but the accuracy was not increased significantly.

The authors Liu, Wang, and Tengfei [28] give a hierarchical attention network in which 2 layers are used to capture semantic information. Then Convolutional structure and Bi-LSTM are combined for the extraction of important meaning to outperform other models. The limitation here is the use of a supervised model and a smaller number of extracted features.

Similarly, Alsubari et al. [29] have used four datasets including hotels, restaurants, Yelp, and Amazon to perform convolutional and max-pooling layers within the CNN are utilized to reduce dimensionality and extract features. Cross-domain analysis has been done but the performance of this model decreases on a single-domain dataset. The work done by Crawford et al. [30] uses inductive transfer learning for the task of hotel review spam detection. But the work is limited to a single domain. As a different approach, Sadiq et al. [31] train deep learning models to predict the star ratings which will match with the review. But prediction of review as fake or real is not done. The work done by Basyar et al. [32] built a Long-Short Term memory model as well as a Gated Recurrent Unit model to detect e-mail spam. The former outperformed the latter, but the result was not significant, and they only used a training dataset due to time constraints.

These deep learning models improve the performance but were considered uninterpretable [33] and perform well on some datasets and badly on others. Models should have consistent performance over different datasets. Deep learning models perform better on large-scale datasets than traditional machine learning models. But deep learning models are computationally expensive. In addition to the above, deep learning models are susceptible to overfitting and thus, they can't be used for smaller data.

Further, Swarm intelligence techniques are also applied in this field. The work done by Shringi et al. [34] use the Fitness-based Grey Wolf Optimization (FGWOK) model and k-means clustering to classify fake and authentic reviews. The datasets used are synthetic spam review (from the Database and Information System Lab, University of Illinois, TripAdvisor dataset), movie review dataset from IMDB, and Yelp review dataset. Upon comparing the findings with different metaheuristic clustering methods like a Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Cuckoo Search (CS), and artificial bee colony clustering, it was found that FGWOK performs better than the current ways.

But the feature interactions have not been considered and better optimizers could have been used.

Similarly, Jacob and Rajendran [35] gives a Fuzzy Artificial Bee colony-based CNN-LSTM approach for fake review classification. After data preprocessing, the chi-squared technique is used for feature extraction and selection and CNNLSTM-FABC is applied, which is found to outperform the earlier approaches. Here also, contextual features are not considered, which may give higher accuracy when combined with the extracted features.

Previously, Saini et al. [36] used k-means artificial bee colony for feature selection after data pre-processing. Then clusters were optimized using ABC with k-means on three datasets namely Synthetic Spam (containing 478 total reviews), Yelp (containing 4952 reviews), and Movie (a subset of IMDB containing 8544 reviews). Earlier Pandey et al. [37] used a combination of cuckoo search and Fermat spiral to identify spam reviews. It is compared to six metaheuristics clustering methods and was found to be better than these six. But still, the accuracy could have been improved further.

Thus, swarm intelligence approaches give an optimized result for the fake review detection problem, but this area has not been explored fully to date. Many bio-inspired algorithms can be utilized and investigated.

Most of the work done till now focuses either on the textual characteristics of the review or the behavior traits of the reviewer. Hence, the former is known as the review-centric model and the latter is known as the reviewer-centric model. The authors, Wang et al. [22] have given a comprehensive fake review detection framework that combines both these models. Yelp and Amazon public datasets have been used to find the effectiveness of their model. But it is limited to the use of only supervised learning models. The authors Kumar et al. [38] devise a novel feature engineering approach named M-SMOTE (Modified-Synthetic Minority Over-sampling TechniquE) model which combines both review-centric and reviewer-centric features. The aim included addressal of the problem of class imbalance. Hence, six reviewer-centric and six review-centric features are extracted. A combination of both these models resulted in higher accuracy. But the work is limited to the above-mentioned features only, which may not apply to other datasets.

Earlier, Budhi et al. [39] found 133 unique features which are a mixture of textual and behavior to detect fake reviews using machine learning methods. This approach was limited to balanced datasets only while in real life most of the datasets are imbalanced and the number of fake and real reviews is different. Two sampling methods were also implemented to improve the accuracy of balanced datasets.

The authors Alawadh et al. [40] perform discourse analysis on the reviews, based on the premise that factual reviews have strong coherence, whereas fraudulent reviews lack structure and semantics. But the dataset used is limited to a smaller review base as compared to the deep learning ones. In other work, the authors Alawadh et al. [41] have given a discourse analysis-based credibility check scheme, which gives higher performance. Use of convolutional neural network with discourse markers as various n-grams, on different proportions of data split reported better results; but only a small, balanced dataset was used, whose size could be increased to produce a scalable solution.

Similarly, Chopra and Dixit [42] investigate Push Ratings (PRs) and Nuke Ratings (NRs) in fake reviews. Push ratings are higher ratings and nuke are lower ratings. Then opinion mining is applied to find whether the review is genuine or not. But use of bag-of-words for preprocessing which results in a sparse matrix, may lead to computational inefficiency.

The reviewer-centric model focuses on the atypical and suspicious behavior of the reviewer as well as identifying the connections between a group of reviewers. These are non-linguistic characteristics of the reviews.

The authors Tufail et al. [15] have extracted the behavioral features of the user. For example, review time, writing style, relationship words, grammatical errors, punctuation, etc. These features may contribute to the classification of the review. This robust model only uses supervised model which may not give apt results for big datasets. Alsubari et al. [17] ponder upon the fact that 90% of genuine reviewers usually write one review a day, based on the product they bought or the service they used. But 70% of fraudsters may write up to 5 reviews a day. But the dataset used is limited to the hotel domain. Hence, the number of reviews per day is an important characteristic of detecting fake reviews [43]. But the number of review features may not apply to other domains. Also, the proportion of positive fake reviews given by the reviewers is more as compared to the proportion of

negative fake reviews [30]. But the work is limited to a single domain. The rating of the fake reviewer tends to be different from the ratings of the genuine reviewer. This can also help in identifying fraudsters [44].

The work done by Abdulqader et al. [5] covers the non-verbal actions or conduct of the reviewer. It is said that usually the fake reviewer posts as soon as possible in an attempt to mislead customers. They either give maximum rating stars or minimum, depending upon the fake positive or negative review. This may be attributed to the algorithms of e-commerce engines that detect fake reviews after a certain period of days and then only remove them. But the damage has already been done till then. One can ascertain the reviewer's reputation by calculating the ratio of the reviewer's friends and followers. In addition to these characteristics, users with longer profile timelines would be more authentic than the users who have recently created their profiles. Even if there is a limited quantity of non-verbal features being employed, the computational cost is less as compared to review-centric feature selection. But this is a purely theoretical model and its practical implementation is yet to be seen.

Another view of the authors Sadiq et al. [31] says that the difference between the ratings given in a review and the emotional intensity of the review is also a giveaway. But prediction of whether the review is fake or real is not there. The authors Hlee et al. [45] collected 4450 reviews from yelp and showed that online reviews of new restaurants are manipulated. The relationship between extreme ratings and fake reviews has been explained. But again, this is just a theoretical model. Tang, Qian, and Yu [46] used a Generative Adversarial Network (GAN) trained model to identify six behavior features including text, rating, and attribute features.

Thus, some of the important reviewer-centric features can be quantity, user profile, timespan, source credibility, non-immediacy, etc. The fake review detection model shows good performance with the incorporation of these non-textual features. This means analyzing reviews using the behavior of the review's author or creator. But which behavioral features are to be selected for the detection problem, is a big task.

Review-centric model of fake review detection focuses on the textual content of the reviews. Research has highlighted that genuine and fraudulent reviews exhibit substantial

linguistic distinctions, which help in identifying them. These features may include micro-linguistic content or semantic content such as product characteristics.

Text features such as number of nouns, verbs, and adjectives along with the use of strong positive and negative words in the review can be identified as a possibility of fake review [17]. The writing style of the reviewer can have lexical characters such as the number of characters and their proportion to uppercase letters, numeric characters, to the rate of spaces or tabs, called stylometric features can help in detecting fake reviews. But use of common features and no introduction of new features for fake review detection is a limitation. The authors Moon et al. [47] obtain fake and authentic reviews of hotels via a survey and determine that features such as lack of details are part of fraudulent reviews. But this work is based on survey and is limited to just the hotel domain. Kumaran et al. [24] have used language features like unigrams and their frequency, bigrams and their frequency and length of reviews but conclude that behavioral features need to be added for accurately identifying the reviews. But focus was more on sentiment classification as positive or negative.

The work done by Abdulqader et al. [5] finds that the short length of the online review, Term Frequency-Inverse Document Frequency (TF-IDF), cohesion and coherence measures, and stylometric features are some of the telltale signs of a spam reviewer. Along with these, LIWC such as less usage of personal pronouns (less use of “I”, “we” etc.), less information about time and location, and strong use of positive as well as negative words, all point to fake reviewers. But it is based on theory, which may or may not apply to other datasets. The authors Wang and Kuan [48] went a step further and divided the features into message level (review length and psychological cues), formulation level (readability and linguistic variables), and control level (rating extremity, review age, etc.). Although the dataset used was Yelp (till 2012), idea was to understand the process human brains undergo to formulate a message. The limitation was that the focused was on single domain and linguistic summary variables rather than composite variables.

According to Alsubari et al. [49] linguistic inquiry and word count is a deep linguistic feature. Its output such as emotions, self-reference, social words, overall cognitive words, etc. can be incorporated to find the fake review. Text similarity is widely used as an

indication of fake reviews because spammers tend to copy the reviews to save effort [25]. But the study of textual and behavioral models separately decreases the accuracy of fake review detection.

Although review-centric features have been used extensively by researchers, it was found that they alone, or reviewer-centric features alone, are not sufficient to determine fake reviews accurately. Thus, a combination of important features of both should be incorporated to make any fake review detection model outperform others. A summary of the research work studied has been given in Appendix Table 1.

3 RESEARCH GAPS

Most of the research work done so far has focused on machine learning and deep learning models. But these models suffer from many limitations.

1. **Class imbalance in datasets -** Several of the datasets are in English only with a huge class imbalance. The majority of the reviews are annotated as authentic as opposed to fake [50] [23]. Statistical or augmentation techniques are used to remove the biasness of trained model. There is a need to investigate the problem of fake review detection from the data imbalance perspective.

2. **Cross Domain solution –** Existing research mostly focuses on a single domain for training and testing. Thus, if a model can be trained in one domain and tested in another, it can also solve the problem of the lack of labeled datasets.

3. **Multilingual fake reviews detection -** There is a substantial lack of research in other languages such as Hindi, Arabic, German, French, etc. This is a challenging area as there is a small amount of data available to train in these languages, as compared to the English language.

4. **Handling Concept Drift problem -** One of the problems is concept drift [21]. Concept drift is that, over a period of time, fake reviewers learn to write reviews similar or comparable to genuine reviewers. Fake review detection should learn to

adapt to differentiate between the two. Sustenance is important to contain the problem of concept drift. According to the authors, there is a strong negative correlation between concept drift and fake review detection models implemented so far, because the reviewers adjust their fake review over time, making the algorithmic solution criteria useless. Thus, concept drift problem must be taken into consideration for a robust solution.

5. **Need for publicly available fake review detection tools -** Google, Amazon, Yelp, etc., all employ their fake review detection algorithms, but these are not available publicly. For example, Amazon detects and deletes fake reviews over a period of time. But till then, most of the damage has already been done. Some websites like fakespot [51] or reviewmeta [52] have been made available for fake review detection, but whether or not they are accurate, is an open-ended question. There is a need for an accurate tool so that users themselves can detect fake reviews using the cues. Also, if the e-commerce websites can identify fake reviews in real time, it will help in increasing the trustworthiness of that e-commerce platform.

6. **Need for novel feature sets –** A limited number of studies have demonstrated the functioning of advanced machine learning and deep learning models using various feature sets, but many limitations have been observed in these. While traditional machine learning models need trained datasets, they also use pre-determined features. There is a need for novel feature sets revolving around reviewer-centric, behavioral, or review-centric parameters. To the best of my knowledge, business-centric features have not been explored so far. When a new business enters the e-commerce market or an existing business launches a new product, it has a higher chance of buying fake reviews. This is done to boost sales. These products might be older products with many reviews and low average ratings also. This may be the business-centric 1characteristic to detect fake reviews. He, Hollenbeck and Prosperio [2] have documented the type of sellers who tend to purchase fake reviews and the timing of their effect on reviews. For example, private Facebook groups bought by Amazon sellers to promote their products and leave a 5-star

review in lieu of a full refund and/or a commission. Table 1 shows a summary of the research gaps found from the literature survey.

2 PROBLEM DEFINITION

This work aims to explore the existing machine learning and deep learning models for fake review detection and develop a framework for efficiently classifying the online reviews as fake or authentic using artificial intelligence models; and validate the accuracy of the proposed framework along with optimizing the time taken.

3 OBJECTIVES OF THIS RESEARCH

- To study and analyze the existing models or frameworks used for fake review detection and evaluate their strengths and weaknesses.
- To handle imbalance in datasets using machine learning
- To design and develop a framework for fake review detection problem by applying various artificial intelligence models including machine-learning and deep-learning networks and identifying reviews as fake or authentic.
- To address the issue of data insufficiency in various unexplored domains via cross-domain application of efficient frameworks.
- To validate the accuracy of the proposed framework for the detection of fake reviews after selecting the most suitable combination of models and optimize the time taken by fine-tuning the designed framework and maximizing accuracy.

4 RESEARCH METHODOLOGY

This research aims to design an efficient framework for fake review detection problem using artificial intelligence models for publicly available fake review datasets. For this, a methodology comprising of 6 phases is proposed: Existing classification model study and comparison, Dataset Acquisition, Data Preprocessing, Applying various machine learning and deep learning models, Comparison of results and accuracy, and Optimization. These phases help in developing an efficient framework for efficiently classifying text reviews in

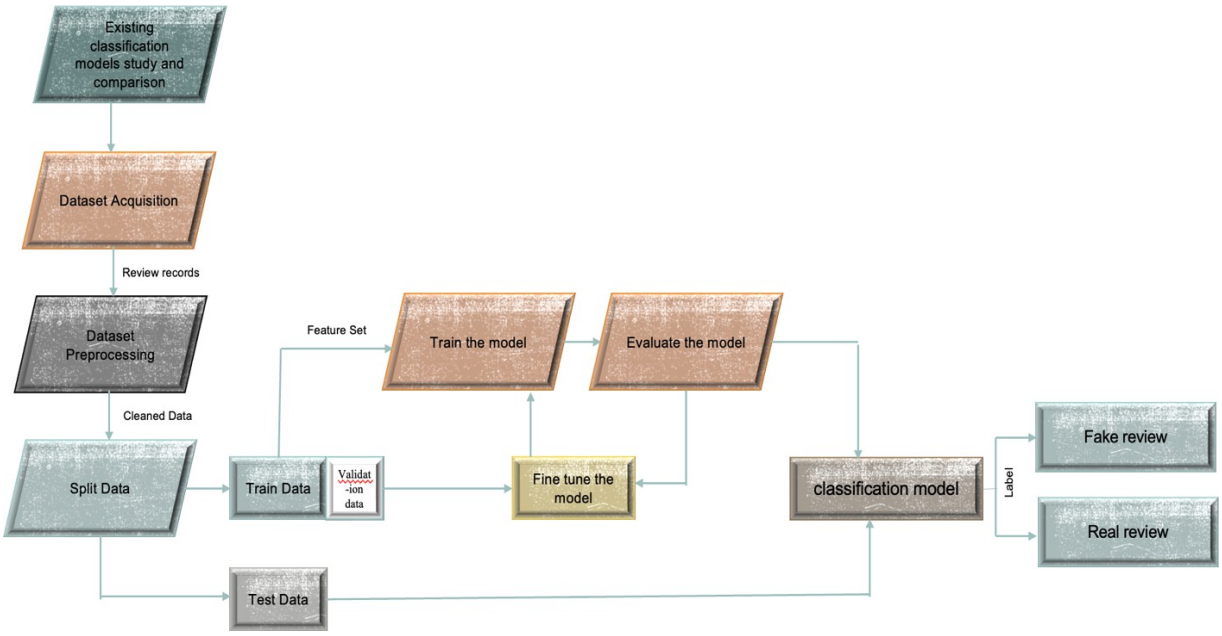
various domains as fake or authentic. Figure 1 shows the flowchart of the proposed methodology.

Phase 1: Existing classification model study and comparison

The first phase of this research involves an extensive literature study of the existing fake review detection models. It will help in developing a deeper understanding of the gaps that exist in the current research and in developing an efficient, better framework for classifying the reviews. Better interpretation and better understanding of heuristic rules will be possible only after comparing the existing classification models.

Moreover, the research gaps will be evaluated after the above study and will be addressed in this research to improve performance, interpretability, scalability, and applicability of classifier models.

Figure 1: Proposed Research Methodology



Phase 2: Dataset Acquisition

This phase involves selecting suitable benchmark dataset(s). Data is a vital component of any AI model, because the result is directly affected by it. The datasets belong to a single

domain such as restaurant, travel, health, online shopping, or services. Initially, a single domain annotated dataset will be used, which may be further increased to multiple domains as the research progresses. The dataset selected should be reputable and of a suitable size. These datasets can be downloaded and used from public domains such as Kaggle, GitHub, Google, amazon e-commerce website, etc.

Phase 3: Data Preprocessing

After selecting a suitable dataset, for the usage of raw data, it needs to be preprocessed to find the meaning of the data or to filter out irrelevant data. In this phase, both fake and authentic reviews are pre-processed to obtain list-of-words. After this, various Natural Language Processing algorithms such as tokenization, stop-word removal, lemmatization, word embeddings, etc. are applied. Punctuations and lowercase will be taken care of.

Apart from the above, for supervised machine learning models, feature extraction and selection is an important step. Deep learning models can extract features internally. Extracting relevant features reduces the dimensionality of the data. This is useful for large datasets because it makes the data manageable and requires less computing time. The models which will be used in the future steps will be trained on these features and then tested. Features can be extracted via algorithms such as Bag-of-words, TF-IDF, etc.

Phase 4: Applying various AI models for fake review classification

Identification of reviews as fake or authentic will be done using Artificial Intelligence models including deep learning and bio-inspired algorithms. This research will be able to differentiate between machine-generated positive and negative reviews and human-created content. The AI model that will be used in this research will be trained on annotated benchmark datasets such as Yelp, Amazon, etc. These corpora contain reviews labeled as 'human-created fake' and 'human-created real' reviews. Using this trained model, machine-created fake reviews can be found. The dataset will be split into training and testing data in various proportions such as 80-20, 70-30, and 60-40. The training dataset will be used for fitting the dataset to our model. The testing dataset will be used to evaluate the fit model.

Initially, some baseline supervised machine learning models will be implemented on the selected database, to represent standard performance. After this, deep learning model(s) and transformers will be chosen, that will outperform the above supervised machine learning models. Parameters will be tuned to further improve the accuracy on the said database for classifying the reviews as fake or authentic. Later, bio-inspired algorithms may be applied to improve accuracy even further. The hyperparameters of the fake review detection must be appropriately selected to improvise the results.

Phase 5: Comparison of result and accuracy

Once the framework for fake review detection problem has been developed, this phase will evaluate the performance and compare the evolution of models applied (from supervised machine-learning to deep-learning to bio-inspired). They will be evaluated on how proficiently it can detect and classify fake reviews from real, using various metrics. The standard metrics include accuracy, precision, recall and the F1-score. These metrics provide a quantitative measure of the accuracy and effectiveness of the classifier, in turn identifying areas for improvement.

Phase 6: Optimization

Once the framework is evaluated, strategies to optimize the said framework will be implemented for timely decision-making and assurance of high quality of research.

5 EXPECTED OUTCOMES OF THIS RESEARCH

This research aims to help consumers and businesses in detecting fake reviews easily, by using a robust, scalable, and optimized framework involving AI models, which will apply to all types of datasets.

References

[1] R. Amos, "Consumer Protection on the Web with Longitudinal Web Crawls and Analysis," 2022.

[2] S. He, B. Hollenbeck and D. Prosperpio, "The Market for Fake Reviews," *Marketing Science,* pp. 896-921, 2022.

[3] M. Zaman, T. Vo-Thanh, C. T. Nguyen, R. Hasan, S. Akter, M. Mariani and L. Hikkerova, "Motives for posting fake reviews: Evidence from a cross-cultural comparison," *Journal of Business Research,* vol. 154, p. 113359, 2023.

[4] J. Salminen, C. Kandpal, A. M. Kamel, S.-g. Jung and B. J. Jansen, "Creating and detecting fake reviews of online products," *Journal of Retailing and Consumer Services ,* vol. 64, p. 102771, 2022.

[5] M. Abdulqader, A. Namoun and Y. Alsaawy, "Fake Online Reviews: A Unified Detection Model Using Deception Theories," *IEEE Access,* vol. 10, pp. 128622-128655, 2022.

[6] Marco-Franco, J. Emilio, P. Pita-Barros, D. Vivas-Orts, S. González-de-Julián and D. Vivas-Consuelo, "COVID-19, fake news, and vaccines: should regulation be implemented?," *International journal of environmental research and public health,* vol. 18, no. 2, p. 744, 2021.

[7] A. Kumar and K. Saroj, "Impact of Customer Review on Social Media Marketing Strategies," *International Journal of Research in Business Studies,* vol. 5, no. 2, pp. 105-114, 2020.

[8] C. Filho, M. D. N. Rafael, L. S. G. Barros and E. Mesquita, "Mind the fake reviews! Protecting consumers from deception through persuasion knowledge acquisition," *Journal of Business Research,* vol. 156, p. 113538, 2023.

[9] M. Kostromitina, D. Keller, M. Cavusoglu and K. Beloin, "“His lack of a mask ruined everything.” Restaurant customer satisfaction during the COVID-19 outbreak: An analysis of Yelp review texts and star-ratings," *International Journal of Hospitality Management,* vol. 98, p. 103048, 2021.

[10] W. Yuanyuan, E. W. Ngai, P. Wu and C. Wu, "Fake online reviews: Literature review, synthesis, and directions for future research," *Decision Support Systems,* vol. 132, p. 113280, 2020.

[11] J. Wang and C. Wu, "Camouflage is NOT easy: Uncovering adversarial fraudsters in large online app review platform," *Measurement and Control 53,* pp. 2137-2145, 2020.

[12] C. Mattson, R. L. Bushardt and A. R. Artino Jr., "When a Measure Becomes a Target, it Ceases to Be a Good Measure.," *Journal of Graduate Medical Education,* vol. 13, no. 1, pp. 2-5, 2021.

[13] E. Ismagilova, E. Slade, N. P. Rana and Y. K. Dwivedi, "The effect of characteristics of source credibility on consumer behaviour: A meta-analysis.," *Journal of Retailing and Consumer Services,* vol. 53, p. 101736, 2020.

[14] L. Floridi and M. Chiriatti, "GPT-3: its nature, scope, limits, and consequences.," *Minds and Machines,* vol. 30, pp. 681-694, 2020.

[15] H. Tufail, M. U. Ashraf, K. Alsubhi and H. M. Aljahdali, "The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection," *IEEE Access, 10,* pp. 25555-25564, 2022.

[16] R. Poonguzhali, S. F. Sowmiya, P. Surendar and M. Vasikaran, "Fake Reviews Detection using Support Vector Machine," in *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, 2022.

[17] S. N. Alsubari, S. N. Deshmukh, A. A. Alqarni, N. Alsharif, T. H. Aldhyani, F. W. Alsaade and O. l. Khalaf, "Data analytics for the identification of fake reviews using supervised learning," *CMC-Computers, Materials & Continua 70,* pp. 3189-3204, 2022.

[18] C. Jing-Yu and W. Ya-Jun, "Semi-Supervised Fake Reviews Detection based on AspamGAN," *Journal of Artificial Intelligence,* vol. 4, no. 1, pp. 17-36, 2022.

[19] G. Stanton and A. A. Irissappane, "GANs for semi-supervised opinion spam detection," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, 2020.

[20] R. Mothukuri, A. Aasritha, K. C. Maramella, K. N. Pokala and G. K. Perumalla, "Fake Review Detection using Unsupervised Learning," in *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, 2022.

[21] R. Mohawesh, S. Tran, R. Ollington and S. Xu, "Analysis of concept drift in fake reviews detection," *Expert Systems with Applications,* vol. 169, p. 114318, 2021.

[22] N. Wang, J. Yang, X. Kong and Y. Gao, "A fake review identification framework considering the suspicion degree of reviews with time burst characteristics," *Expert systems with applications,* vol. 190, p. 116207, 2022.

[23] S. Kaddoura, G. Chandrasekaran, D. E. Popescu and J. H. Duraisamy, "A systematic literature review on spam content detection and classification," *PeerJ Computer Science,* vol. 8, p. e830, 2022.

[24] N. Kumaran, C. H. Chowdhary and D. Sreekavya, "Detection of fake online reviews using semi supervised and supervised learning.," *International Research Journal of Engineering and Technology(IRJET),* vol. 8, pp. 650-656, 2021.

[25] N. Hussain, H. T. Mirza, I. Hussain, F. Iqbal and I. Memon, "Spam review detection using the linguistic and spammer behavioral methods," *IEEE Access 8,* pp. 53801-53816, 2020.

[26] X. Ren, Z. Yuan and J. Huang, "Research on Fake Reviews Detection Based on Graph Neural Network," *International Symposium on Computer Applications and Information Systems (ISCAIS 2022),* pp. 290-297, 2022.

[27] Z. Gyongyi, H. Garcia-Molina and J. Pederson, "Combating web spam with trustrank," in *Proceedings of the 30th international conference on very large data bases (VLDB)*, 2004.

[28] Y. Liu, L. Wang, T. Shi and J. Li, "Detection of spam reviews through a hierarchical attention architecture with N-gram CNN and Bi-LSTM," *Information Systems 103,* vol. 103, p. 101865, 2022.

[29] N. S. Alsubari, S. N. Deshmukh, M. H. Al-Adhaileh, F. W. Alsaade and T. H. Aldhyani, "Development of Integrated Neural Network Model for Identification of Fake Reviews in E-Commerce Using Multidomain Datasets," *Applied Bionics and Biomechanics,* 2021.

[30] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter and H. Al Najada, "Using Inductive Transfer Learning to Improve Hotel Review Spam Detection," in *2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI)*, 2021.

[31] S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara and M. Nappi, "Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning," *Expert Systems with Applications,* vol. 181, p. 115111, 2021.

[32] I. Basyar, M. D. Adiwijaya and D. Murdiyansah, "Email spam classification using gated recurrent unit and long short-term memory," *Journal of Computer Science,* pp. 559-567, 2020.

[33] R. Mohawesh, S. Xu, S. N. Tran, R. Ollington, M. Springer, Y. Jararweh and S. Maqsood, "Fake reviews detection: A survey," *IEEE Access,* vol. 9, pp. 65771-65802, 2021.

[34] S. Shringi, H. Sharma and D. L. Suthar, "Fitness-Based Grey Wolf Optimizer Clustering Method for Spam Review Detection," *Mathematical Problems in Engineering,* vol. 2022, 2022.

[35] M. S. Jacob and P. S. Rajendran, "Fuzzy artificial bee colony-based CNN-LSTM and semantic feature for fake product review classification," *Concurrency and Computation: Practice and Experience,* vol. 34, no. 1, p. e6539, 2022.

[36] P. Saini, S. Shringi, N. Sharma and H. Sharma, "Spam Review Detection Using K-Means Artificial Bee Colony," *Communication and Intelligent Systems,* pp. 731-744, 2021.

[37] A. C. Pandey and D. S. Rajpoot, "Spam review detection using spiral cuckoo search clustering method," *Evolutionary Intelligence,* vol. 12, no. 2, pp. 147-164, 2019.

[38] A. Kumar, R. D. Gopal, R. Shankar and K. H. Tan, "Fraudulent review detection model focusing on emotional expressions and explicit aspects: investigating the potential of feature engineering," *Decision Support Systems,* vol. 155, p. 113728, 2022.

[39] G. S. Budhi, R. Chiong, Z. Wang and S. Dhakal, "Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews," *Electronic Commerce Research and Applications,* vol. 47, p. 101048, 2021.

[40] H. M. Alawadh, A. Alabrah, T. Meraj and H. T. Rauf, "Semantic Features-Based Discourse Analysis Using Deceptiveand Real Text Reviews," *Information,* vol. 14, no. 1, p. 34, 2023.

[41] H. M. Alawadh, A. Alabarah, T. Meraj and H. T. Rauf, "Discourse analysis based credibility checks to online reviews using deep learning based discourse markers," *Computer Speech and Language,* vol. 78, p. 101450, 2023.

[42] A. B. Chopra and V. S. Dixit, "Detecting biased user-product ratings foronline products using opinion mining," *Journal of Intelligent Systems,* vol. 32, no. 1, p. 20229030, 2023.

[43] A. Heydari, M. A. Tavakoli, M. N. Salim and Z. Heydari, "Detection of review spam: A survey," *Expert Systems with Applications,* vol. 42, no. 7, pp. 3634-3642, 2015.

[44] M. Ott, C. Cardie and J. T. Hancock, "Negative deceptive opinion spam," in *Proc. of the 2013Conf. of the North American Chapter of the Association for Computational Linguistics: Human LanguageTechnologies*, Atlanta, Georgia, 2013.

[45] S. Hlee, H. Lee, C. Koo and N. Chung, "Fake Reviews or Not: Exploring the relationship between time trend and online restaurant reviews," *Telematics and Informatics,* vol. 59, p. 101560, 2021.

[46] X. Tang, T. Qian and Z. You, "Generating behavior features for cold-start spam review detection with adversarial learning," *Information Sciences,* vol. 526, pp. 274-288, 2020.

[47] S. Moon, M.-Y. Kim and D. Lacobucci, "Content analysis of fake consumer reviews by survey-based text categorization," *International Journal of Research in Marketing,* vol. 38, no. 2, pp. 343-364, 2021.

[48] B. Wang and K. K. Kuan, "Understanding the Message and Formulation of Fake Online Reviews: A Language-production Model Perspective," *AIS Transactions on Human-Computer Interaction,* vol. 14, no. 2, pp. 207-229, 2022.

[49] S. N. Alsubari, M. B. Shelke and S. N. Deshmukh, "Fake reviews identification based on deep computational linguistic," *International Journal of Advanced Science and Technology 29,* pp. 3846-3856, 2020.

[50] Y. Jianrong, Y. Zheng and H. Jiang, "An ensemble model for fake online review detection based on data resampling, feature pruning, and parameter optimization," *IEEE Access 9,* pp. 16914-16927, 2021.

[51] "https://www.fakespot.com/," 2023. [Online]. Available: https://www.fakespot.com/. [Accessed 8 January 2023].

[52] "https://reviewmeta.com/," 2023. [Online]. Available: https://reviewmeta.com/. [Accessed 9 January 2023].

[53] S. Banerjee, "Exaggeration in fake vs. authentic online reviews for luxury and budget hotels.," *International journal of information management,* vol. 62, p. 102416.

[54] Z. Wang, W. Wei, X.-L. Mao, G. Guo, P. Zhou and S. Jiang, "User-based network embedding for opinion spammer detection," *Pattern Recognition,* vol. 125, p. 108512, 2022.

[55] S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," in *Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining*, 2015.

[56] A. Tuomi, "Deepfake consumer reviews in tourism: Preliminary findings," *Annals of Tourism Research Empirical Insights,* 2021.

[57] A. Staff, "Amazon targets fake review fraudsters on social media," 19 July 2022. [Online]. Available: https://www.aboutamazon.com/news/policy-news-views/amazon-targets-fake-review-fraudsters-on-social-media. [Accessed 28 February 2023].

[58] D. Pritchett, "A look at how we tackle fake and fraudulent contributed content," 18 February 2021. [Online]. Available: https://blog.google/products/maps/google-maps-101-how-we-tackle-fake-and-fraudulent-contributed-content/. [Accessed 28 February 2023].

[59] R. Hassan and M. R. Islam, "Detection of fake online reviews using semi-supervised and supervised learning.," in *2019 International conference on electrical, computer and communication engineering (ECCE)*, 2019.

[60] A. Mewada and R. K. Dewang, "Research on false review detection Methods: A state-of-the-art review," *Journal of King Saud University –Computer and Information Sciences,* vol. 34, no. 9, pp. 7530-7546, 2021.

[61] D. Plotkina, A. Munzel and J. Pallud, "Illusions of truth—experimental insights into human and algorithmic detections of fake online reviews," *Journal of Business Research,* vol. 109, pp. 511-523, 2020.

[62] H. Tang and H. Cao, "A review of research on detection of fake commodity reviews," *Journal of Physics : conference series,* vol. 1651, no. 1, p. 012055, 2020.

[63] H. Li, Z. Chen, B. Liu, X. Wei and J. Shao, "Spotting fake reviews via collective positive-unlabeled learning," in *2014 IEEE international conference on data mining*, 2014.

Appendix

Table 1: Summary of research done in fake review detection.

S.No.	Research Publication	Methodology	Research Contribution
1.	Zaman, M., Vo-Thanh, T., Nguyen, C. T., Hasan, R., Akter, S., Mariani, M., & Hikkerova, L. (2023). Motives for posting fake reviews: Evidence from a cross-cultural comparison. *Journal of Business Research, 154*, 113359. [3] (SCIE, Scopus indexed)	• Finds the motives of writing fake reviews. • Focuses on cosmetic industry and French-Korean cultures.	• Focused on customer's perspective. • Used single domain of cosmetics • Gender specific survey
2.	Filho, C., Rafael, M. D., Barros, L. S., & Mesquita, E. (2023). Mind the fake reviews! Protecting consumers from deception through persuasion knowledge acquisition. *Journal of Business Research, 156*, 113538. [8] (SCIE, SCOPUS indexed)	• Claims to help consumers identify fake and real reviews through persuasion knowledge acquisition.	• Theoretical model • Fake and real reviews are mostly comparable. • Found that detection, despite the knowledge acquired, is not easy.
3.	Alawadh, H. M., Alabarah, A., Meraj, T., & Rauf, H. T. (2023). Discourse analysis-based credibility checks to online reviews using deep learning-based discourse markers. *Computer Speech and Language, 78*, 101450. [41] (SCIE, SCOPUS indexed)	• Given a discourse analysis-based credibility check scheme, which gives higher performance. • Used Convolutional Neural Network with discourse markers as various n-grams, on different proportions of data split.	• Discourse analysis has been used • Dataset's size could be increased. • Numbers of channels used in the CNN is less. • Imbalance of dataset not considered.
4.	Chopra, A. B., & Dixit, V. S. (2023). Detecting biased user-product ratings for online products using opinion mining. *Journal of Intelligent Systems, 32*(1), 20229030. [42] (SCIE, SCOPUS indexed)	• Investigate Push Ratings (PRs) and Nuke Ratings (NRs) in the fake reviews. • Apply opinion mining to find whether the review is genuine or not.	• Uses bag-of-words for preprocessing which results in sparse matrix, which may lead to computational inefficiency.

5.	Banerjee, S., 2022. Exaggeration in fake vs. authentic online reviews for luxury and budget hotels. International journal of information management, 62, p.102416. [53] (SCIE, SCOPUS indexed)	• Detects textual cues. • Finds that fake reviews are not always more exaggerated than real reviews.	• The fake reviews created by this research were not written by professionals. • Domain luxury and budget hotels only
6.	Tufail, H., Ashraf, M. U., Alsubhi, K., & Aljahdali, H. M. (2022). The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection. *IEEE Access, 10*, 25555-25564. [15] (SCIE, SCOPUS indexed)	• Used SVM, K-Nearest Neighbor and Linear Regression on Yelp hotel review dataset.	• Not much work has been done on unlabeled datasets. • Applicable on a limited domain
7.	Alsubari, S., Deshmukh, S. N., Alqarni, A., Alsharif, N., Aldhyani, T., Alsaade, F. W., & Khalaf, O. I. (2022). Data analytics for the identification of fake reviews using supervised learning. *CMC-Computers, Materials & Continua 70*, 3189-3204. [17] (SCIE, SCOPUS indexed)	• Applies SVM, Naïve-Bayes, Random Forest, and Adaptive Boost on the extracted features of hotel database.	• Fewer number of extracted features • Dataset limited to hotel domain.
8.	Abdulqader, M., Namoun, A., & Alsaawy, Y. (2022). Fake Online Reviews: A Unified Detection Model Using Deception Theories. *IEEE Access, 10*, 128622-128655. [5] (SCIE, SCOPUS indexed)	• Justify that non-verbal features hold more weight than verbal features and a combination of both can improve the detection accuracy.	• Theory-based model • May not apply to all datasets.
9.	Wang, N., Yang, J., Kong, X., & Gao, Y. (2022). A fake review identification framework considering the suspicion degree of reviews with time burst characteristics. *Expert systems with applications, 190*, 116207. [22] (SCIE, SCOPUS indexed)	• Incorporated time-burst characteristics for fake review detection	• Not applicable on large datasets • Geographical factors not considered in feature extraction.
10.	Kaddoura, S., Chandrasekaran, G., Popescu, D. E., & Duraisamy, J. H. (2022). A systematic literature review on spam content detection and classification. PeerJ Computer Science, 8, e830. [23] (SCIE, SCOPUS indexed)	• Give the drawback of machine learning algorithms. • Manual extraction of features • Need a large number of computational resources.	• A lot of work has been done in the area of machine learning. Further research can be done using deep learning or bio-inspired learning.
11.	Shringi, S., Sharma, H., & Suthar, D. (2022). Fitness-Based Grey Wolf	• Use the Fitness-based Grey Wolf Optimization (FGWOK) technique and K-Means	• Have not taken into account the feature interactions.

	Optimizer Clustering Method for Spam Review Detection. *Mathematical Problems in Engineering, 2022*. [34] (SCIE, SCOPUS indexed)	Clustering to classify fake and authentic reviews. • The datasets used are synthetic spam review, movie review dataset from IMDB and Yelp review dataset	• Better optimizers can be proposed.
12.	Jacob, M. S., & Rajendran, P. (2022). Fuzzy artificial bee colony-based CNN-LSTM and semantic feature for fake product review classification. Concurrency and Computation: Practice and Experience, 34(1), e6539. [35] (SCIE, SCOPUS indexed)	• Gave a Fuzzy Artificial Bee colony-based CNN-LSTM approach for fake review classification.	• Contextual features are not considered. • Higher accuracy may result from contextual features.
13.	Kumar, A., Gopal, R. D., Shankar, R., & Tan, K. H. (2022). Fraudulent review detection model focusing on emotional expressions and explicit aspects: investigating the potential of feature engineering. Decision Support Systems, 155, 113728. [38] (SCIE, SCOPUS indexed)	• Devised an innovative feature engineering approach called M-SMOTE (Modified- Synthetic Minority Over-sampling Technique) model combining review- and reviewer-centric features. • They also try to solve the class imbalance problem.	• Their work is limited to six features only, which may not be applicable to other datasets.
14.	Z. Wang, W. Wei, X.-L. Mao, G. Guo, P. Zhou and S. Jiang, "User-based network embedding for opinion spammer detection.," Pattern Recognition, vol. 125, p. 108512, 2022. [54] (SCI, SCOPUS indexed)	• Amazon dataset • Product review subset taken from Kaggle containing 5,68,454 reviews. • Subset containing cell phone reviews contains 1,039,833 reviews. • Subset containing reviews on 3 categories – baby, musical instruments and automotives.	• Imbalance, • Lack of temporal or metadata, • Age of the dataset
15.	Alsubari, N. S., Deshmukh, S. N., Al-Adhaileh, M. H., Alsaade, F. W., & Aldhyani, T. H. (2021). Development of Integrated Neural Network Model for Identification of Fake Reviews in E-Commerce Using Multidomain Datasets. *Applied Bionics and Biomechanics*. [29] (SCIE, SCOPUS indexed)	• Have used four datasets including hotels, restaurants, Yelp and Amazon to perform convolutional and max-pooling layers of the CNN for dimensionality reduction and feature extraction.	• Good for cross-domain analysis but decreased accuracy on single domain dataset.
16.	Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning. *Expert Systems with Applications, 181*, 115111 [31]	• Train deep learning algorithms to classify the star ratings which will match with the review.	• Not much work done in this area (combination of review and corresponding rating)

	(SCIE, SCOPUS indexed)		
17.	Mohawesh, R., Xu, S., Tran, S. N., Ollington, R., Springer, M., Jararweh, Y., & Maqsood, S. (2021). Fake reviews detection: A survey. IEEE Access, 9, 65771-65802. [33] (SCIE, SCOPUS indexed)	• Compares various techniques. • Deduce that deep learning models are uninterpretable.	• After extensive survey, it was found that there is a need of explainable techniques to detect the fake review.
18.	Mohawesh, R., Tran, S., Ollington, R., & Xu, S. (2021). Analysis of concept drift in fake reviews detection. Expert Systems with Applications, 169, 114318. [21] (SCIE, SCOPUS indexed)	• Assessed the correlation between concept drift and FRD problem and found they are negatively correlated, using 4 datasets	• Efficient method needed to handle the problem of concept drift.
19.	Budhi, G. S., Chiong, R., Wang, Z., & Dhakal, S. (2021). Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews. *Electronic Commerce Research and Applications, 47*, 101048. [39] (SCIE, SCOPUS indexed)	• Found 133 unique features which are a mixture of textual and behaviour to detect fake reviews using machine learning methods.	• Multiple approaches of feature extraction can be used for better accuracy.
20.	Hlee, S., Lee, H., Koo, C., & Chung, N. (2021). Fake Reviews or Not: Exploring the relationship between time trend and online restaurant reviews. *Telematics and Informatics, 59*, 101560. [45] (SCIE, SCOPUS indexed)	• They explained the relationship between extreme ratings and fake reviews.	• This is just a theoretical model, practical implementation yet to be done.
21.	Moon, S., Kim, M.-Y., & Lacobucci, D. (2021). Content analysis of fake consumer reviews by survey-based text categorization. International Journal of Research in Marketing, 38(2), 343-364. [47] (SCIE, SCOPUS indexed)	• Obtain fake and authentic reviews of hotels via a survey and determine the features. • Found that fake reviews tend to have more positive and more negative words	• Their work is based on manual survey and is limited to just the hotel domain
22.	Hussain, N., Mirza, H. T., Hussain, I., Iqbal, F., & Memon, I. (2020). Spam review detection using the linguistic and spammer behavioral methods. IEEE Access 8, 53801-53816. [25] (SCIE, SCOPUS indexed)	• Given some important behavioural and linguistic pointers to identify fake reviews.	• Decreased accuracy due to two separate models on behavioral and linguistic features. • Features may be combined to improve accuracy.
23.	Basyar, I., Adiwijaya, M. D., & Murdiyansah, D. (2020). Email spam classification using gated recurrent unit and long short-term memory. *Journal of Computer Science*, 559-567. [32] (SCOPUS indexed)	• Built LSTM and GRU to detect email spam.	• Result is not significant because they focused only on training dataset.
	Alsubari, S. N., Shelke, M. B., & Deshmukh, S. N. (2020). Fake reviews	• Uses deep linguistic features such as linguistic enquiry and	• Behavioural data not considered.

24.	identification based on deep computational linguistic. *International Journal of Advanced Science and Technology 29*, 3846-3856. [49] (SCIE, SCOPUS indexed)	word count to find the fake review.	
25.	S. Rayana and L. Akoglu, "Collective opinion spam detection: Bridging review networks and metadata," in Proceedings of the 21th acm sig kdd international conference on knowledge discovery and data mining, 2015. [55]	• Yelp Hotel and restaurant datasets - YelpChi, YelpNYC, YelpZip • The YelpChi dataset includes a subset of reviews and associated data from Yelp's platform, specifically pertaining to businesses in the Chicago area. • The YelpNYC dataset comprises a subset of reviews and related information from businesses located in the New York City area. • The YelpZip dataset encompasses reviews and associated data from various zip codes, providing a more diverse sample of businesses and user reviews.	• Imbalance in dataset • Dataset specificity
26.	L. Huayi, Z. Chen, B. Liu, X. Wei, and J. Shao, "Spotting fake reviews via collective positive-unlabeled learning," in 2014 IEEE international conference on data mining, 2014. [30]	• Multi domain deception dataset • Another benchmark dataset of Hotels, Restaurants, and doctors review (2,836 reviews). Hotel review comprise of 1880 reviews, Restaurants 400, and doctor 556 reviews	• Lack of metadata, • Labeling accuracy, limited size
27.	M. Ott, C. Cardie, and J. T. Hancock, "Negative deceptive opinion spam," in Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies, 2013. [44]	• Single domain hotel review dataset • Benchmark dataset consisting of 1,600 hotel reviews (800 spam and 800 ham) from TripAdvisor website.	• Limited size of dataset, • Single Annotation, • crowdsourced spam, dataset specificity

MANAV RACHNA
Vidyanatariksha

Dr. O.P. Bhalla Central Library
Manav Rachna International Institute of Research and Studies
(A deemed-to-be-university under section 3 of UGC Act 1956)

SIMILARITY VERIFICATION CERTIFICATE

Title: Developing a framework for fake review detection using Deep learning techniques

Researcher: Richa Gupta
Registration Number: 21/Ph·D/108

Supervisor: Dr. Ankur Kashyap
Department: CSE

This is to certify that the above mentioned Short Synopsis/Thesis was scanned for similarity index through Turnitin Software on 22/11/23 with submission ID 2235689236 and the similarity index was found to be 6 %. The complete report running into 34 Pages, 8078 Word count with File Size 5420 KB is being forwarded to the Research Supervisor/DRC/DRB for review.

Date: 23/11/23

Dr. Rajesh Kumar 23/11/23
University Librarian
Dr. O.P. Bhalla Central Library
Manav Rachna International
Institute of Research & Studies
Faridabad

The complete report of the above Short Synopsis/Thesis has been reviewed and the similarity index is below the accepted norms as per the University Grants Commission (Promotion of Academic Integrity and Prevention of Plagiarism in Higher Educational Institutions) Regulations 2018. The Short Synopsis/Thesis may be considered for further processing

Research Scholar

23/11
Research Supervisor

DEVELOPING A FRAMEWORK FOR FAKE REVIEW DETECTION USING DEEP LEARNING TECHNIQUES

by Richa Gupta

Submission date: 22-Nov-2023 07:02AM (UTC+0530)
Submission ID: 2235689236
File name: RK_FOR_FAKE_REVIEW_DETECTION_USING_DEEP_LEARNING_TECHNIQUES.docx (1.88M)
Word count: 8078
Character count: 44538

DEVELOPING A FRAMEWORK FOR FAKE REVIEW DETECTION USING DEEP LEARNING TECHNIQUES

SHORT SYNOPSIS

By

Ms. Richa Gupta

Reg. No.: **21/Ph.D/108**

Under the supervision of

Supervisor

Dr. Indu Kashyap
Professor
Computer Science and Engineering
School of Engineering & Technology
MRIIRS

Joint Supervisor

Dr. Vinita Jindal
Professor
Dept of Computer Science
Keshav Mahavidyalaya
University of Delhi

School of Engineering and Technology

MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH & STUDIES

SECTOR-43, SURAJKUND-DELHI ROAD, FARIDABAD - 121001

DEVELOPING A FRAMEWORK FOR FAKE REVIEW DETECTION USING DEEP LEARNING TECHNIQUES

ORIGINALITY REPORT

Dr. Rajesh Kumar
University Librarian
Bhalla Central Library
Manav Rachna International
Faridabad

6%	4%	6%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.journaltocs.ac.uk Internet Source	1%
2	www.hindawi.com Internet Source	1%
3	"Communication and Intelligent Systems", Springer Science and Business Media LLC, 2021 Publication	<1%
4	shodhgangotri.inflibnet.ac.in Internet Source	<1%
5	Ajay Kumar, Ram D. Gopal, Ravi Shankar, Kim Hua Tan. "Fraudulent review detection model focusing on emotional expressions and explicit aspects: investigating the potential of feature engineering", Decision Support Systems, 2022 Publication	<1%
6	www.semanticscholar.org Internet Source	<1%

7	Sushil Kumar Maurya, Dinesh Singh, Ashish Kumar Maurya. "Deceptive opinion spam detection approaches: a literature survey", Applied Intelligence, 2022 Publication	<1 %
8	thesai.org Internet Source	<1 %
9	Jianrong Yao, Yuan Zheng, Hui Jiang, Lu Wang. "An ensemble model for fake online review detection based on data resampling, feature pruning, and parameter optimization", IEEE Access, 2021 Publication	<1 %
10	www.researchgate.net Internet Source	<1 %
11	Joni Salminen, Chandrashekhar Kandpal, Ahmed Mohamed Kamel, Soon-gyo Jung, Bernard J. Jansen. "Creating and detecting fake reviews of online products", Journal of Retailing and Consumer Services, 2022 Publication	<1 %
12	Ning Wang, Jun Yang, Xuefeng Kong, Ying Gao. "A fake review identification framework considering the suspicion degree of reviews with time burst characteristics", Expert Systems with Applications, 2022 Publication	<1 %

	dspace.bracu.ac.bd:8080 Internet Source	<1 %
14	keshav.du.ac.in Internet Source	<1 %
15	opus.bibliothek.uni-augsburg.de Internet Source	<1 %
16	Shudong Li, Guojin Zhong, Yanlin Jin, Xiaobo Wu, Peican Zhu, Zhen Wang. "A Deceptive Reviews Detection Method Based on Multidimensional Feature Construction and Ensemble Feature Selection", IEEE Transactions on Computational Social Systems, 2022 Publication	<1 %
17	Avinash Chandra Pandey, Dharmveer Singh Rajpoot. "Spam review detection using spiral cuckoo search clustering method", Evolutionary Intelligence, 2019 Publication	<1 %
18	mpra.ub.uni-muenchen.de Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 14 words
Exclude bibliography	On		